基于概率密度 PCA 的多模态过程故障检测 *

郭金玉，刘玉超，李　元

(沈阳化工大学 信息工程学院，沈阳 110142)

摘　要：为了提高故障检测和分类能力，提出基于概率密度 PCA 的多模态过程故障检测算法。对各模态的训练数据建立 PCA 模型，计算各个模型的控制限和匹配系数。根据匹配系数计算各模态统一的控制限。对新来的数据，运用概率密度确定其模态。新来数据向对应模态的模型上投影并计算统一的统计量，比较统计量与控制限进行多模态过程故障检测。把该方法应用到数值例子和半导体过程中，仿真结果表明，该算法在分类及多模态过程故障检测方面具有很高的准确性。

关键词：多模态过程；故障检测；概率密度；主元分析；匹配系数

中图分类号：TP277　　doi: 10.3969/j.issn.1001-3695.2017.12.0838

PCA based on probability density for fault detection of multimodal processes

Guo Jinyu, Liu Yuchao, Li Yuan

(*College of Information Engineering*, *Shenyang University of Chemical Technology*, *Shenyang* 110142, *China*)

Abstract: In order to improve the ability of fault detection and classification, this paper proposed PCA based on probability density for fault detection of multimodal processes. It established PCA model for training data of each mode, and calculated the control limits and matching coefficients of each model. It calculated the unified control limit of each mode according to the matching coefficients. For a new data, it determined its mode by the probability density. It projected the new data to PCA model of the corresponding mode and calculated the unified statistics. It performed fault detection of multimodal processes by comparing the statistics with control limit. We applied the method to a numerical example and the semiconductor process. Simulation results show that the proposed algorithm has high accuracy in classification and fault detection of multimodal processes.

Key words: multimodal processes; fault detection; probability density; principal component analysis(PCA); matching coefficients

0　引言

近年来，随着科技和自动化水平的不断提高，工业生产过程越来越复杂，这使得过程监控越来越受到人们的重视[1~3]。其中基于数据驱动的故障检测技术受到国内外学术界的广泛关注，以主元分析（principal component analysis, PCA）算法为代表的多元统计分析方法迅速发展，并衍生出多种新的故障检测方法[4~6]。PCA 是一种对数据进行化简的方法，但是它只能提取数据的全局信息，会造成局部信息的丢失，而且处理非线性问题效果不理想。

多模态间歇过程具有严重的非线性、时变性和多工况等特性，这使得多模态间歇过程的故障诊断更加具有挑战性。近年来，许多学者通过不同角度分析多模态工业过程，并提出了多种故障诊断方法[7,8]。为了解决具有非高斯、非线性的多模态间歇过程故障检测问题，He 等人[9]提出基于 k 近邻(k-nearest neighbor, KNN)的故障检测算法。KNN 算法对多模态非线性问题是有效的，但是对于样本容量大的数据集其计算量比较大，而且当多模态样本方差差别较大时，检测效果不理想。另外，KNN 算法每次分类都会重新进行全局运算。为了建立单个模型来实现多模态过程的监控目的，马贺贺等人[10]提出基于局部离群因子(local outlier factor，LOF)的方法进行故障检测。因为 LOF 值适用于基于不同密度的数据集群，此时可以得出此对象的 LOF 值，以更确切地判断其是否为一个噪声点。但此方法也具有一定的局限性，并不能适用于所有的场景。此外，它

收稿日期：2017-12-05；**修回日期：**2018-01-23　　**基金项目：**国家自然科学基金重大项目（61490701）；国家自然科学基金资助项目（61174119）；辽宁省教育厅重点实验室项目（LZ2015059）；辽宁省自然科学基金资助项目（201602584）；辽宁省教育厅项目（L2016007，L2015432）

作者简介：郭金玉（1975-），女，山东高唐人，副教授，主要研究方向为工业过程的故障诊断、生物特征识别算法及应用（shandong401@sina.com）；刘玉超（1994-），女，辽宁朝阳人，硕士研究生，主要研究方向为工业过程的监视、故障检测和诊断；李元（1964-），女，辽宁沈阳人，教授，博士，主要研究方向为基于数据驱动的复杂工业过程故障诊断.

的计算也相对较为复杂，准确性不高。最近，为了能及时检测多工况生产过程中的故障，刘帮莉等人[11]提出基于局部密度估计的多模态过程监控策略。针对间歇过程的多工况和非线性特征，李元等人[12]提出基于动态多向局部离群因子算法（dynamic multiway local outlier factor，DMLOF）。Deng 等人[13]提出基于局部近邻相似度分析的多模态故障检测方法，将该方法应用到连续过程的故障检测中。郭金玉等人[14]在此基础上，提出基于在线升级主样本建模的批次过程 KNN 故障检测方法。本文提出一种基于概率密度 PCA（PCA based on probability density, PD-PCA）的多模态间歇过程故障检测算法。运用概率密度函数[15]进行模态识别，在此基础上，运用 PCA 进行多模态间歇过程故障检测。该方法可以有效地判断数据的模态，使得多模态间歇过程故障诊断结果更加准确。

1　基于概率密度 PCA 的多模态过程故障检测

1.1　概率密度函数

核密度估计（Kernel density estimation，KDE）[16,17]，是一种用于估计概率密度函数的非参数方法。它的基本思想是用假定的正态分布计算分布中统计样本的密度。设 $x_1,x_2,\cdots x_n$ 是服从独立同分布的随机变量，其服从的分布密度函数为 $f_h(x)$，$x\in R$，定义函数：

$$\hat{f}_h(x)=\frac{1}{nh}\sum_{j=1}^{n}K\{\frac{x-x_i}{h}\} \tag{1}$$

为密度函数 $f_h(x)$ 的核密度估计。其中：$K(?)$ 称为核函数；h 为预先给定的正数，通常称为窗口宽度或光滑参数；n 为样本个数。概率密度函数用于检测当前时刻数据点属于哪个参考训练数据集，较大的密度表示当前时刻的数据在训练数据集内，本文利用该属性进行模态识别。对于一个新获得的样本 $\boldsymbol{x}_{new}$，其核密度估计器为

$$f_i(x_{new})=\frac{1}{mh}\sum_{j=1}^{m}K\{\frac{x_{new,j}-x_{i,j}}{h}\} \tag{2}$$

其中：$i=1,2,\cdots,n$；$\boldsymbol{x}_{new,j}$ 表示 $\boldsymbol{x}_{new}$ 的第 j 个变量值；$\boldsymbol{x}_{i,j}$ 表示训练数据集中第 i 个样本的第 j 个变量值；h 是窗口宽度；m 为变量个数。核函数 K 决定了函数图形的凸起程度，目前存在很多的概率密度函数。在实际应用中，高斯核函数是应用最广泛的。本文采用高斯核，其核估计函数为

$$f_i(x_{new})=\frac{1}{m}\frac{1}{h\sqrt{2\pi}}\sum_{j=1}^{m}\exp\left[-\frac{(x_{new,j}-x_{i,j})^2}{2h^2}\right] \tag{3}$$

关于窗口宽度 h 的选择，已有大量文献研究，建议根据具体情况选择合适的经验值。正常状态下的训练数据集容易获得有效的信息，因此可以保证 KDE 的性能。如果当前样本与训练样本同属于一个模态，其密度值通常会比较大；否则，密度值将很小，接近于零。

1.2　多模态 PCA 模型的建立

由于在多变量统计分析中，主元分析可以有效地提取数据特征，并对原始数据起到降低维数的作用[18]，所以对多个操作模式的过程进行监控时需要建立多个 PCA 模型[19]。假设生产过程包含 N 个操作模态，$X_r^{(s)}$ (s=1,2…N) 是某一模态矩阵，其中 r 为该模态的标签，对 $X_r^{(s)}$ 进行主成分分解为

$$X_r^{(s)}=\hat{X}_r^{(s)}+E=T^{(s)}P^{(s)T}+E \tag{4}$$

其中：$T^{(s)}$ 是主元得分矩阵；$P^{(s)}$ 是负载矩阵；E 是残差向量。在利用 PCA 进行故障检测时，常用的统计指标有平方预测误差（squared prediction error, SPE）（也称 Q 统计量）和 Hotelling's T^2。SPE 统计量反映了输入变化量的测量值对主元模型的偏离程度，是对模型外部变化的一种度量；T^2 统计量反映了每个主元在变化趋势和幅值上偏离模型的程度，是对模型内部变化的一种度量，它可以用来对多个主元同时检测。

模态 r 的 SPE 统计量为

$$SPE_{r,\alpha}=\left\|(I-PP^Tx_r\right\|^2\le\delta_\alpha^2 \tag{5}$$

其中：δ_α^2 表示置信水平为 α 的控制限，当 SPE 在控制限内时，即认为是正常样本。δ_α^2 控制限的计算公式为

$$\delta_\alpha^2=\theta_1(\frac{C_\alpha\sqrt{2\theta_2h_0^2}}{\theta_1}+1+\frac{\theta_2h_0(h_0-l)}{\theta_1^2})^{1/h_0} \tag{6}$$

其中：$\theta_i=\sum_{j=l+1}^{m}\lambda_j^i$（$i$=1, 2, 3）；$h_0=1-2\theta_1\theta_3/3\theta_1^2$；$c_\alpha$ 为标准正态分布在置信水平 α 下的阈值；l 为 PCA 模型的主元个数；m 为矩阵 X 的维数。

模态 r 的 Hotelling's T_r^2 统计量定义为

$$T_{r,\alpha}^2=x_r^TP\Lambda^{-1}P^Tx_r\le T_\alpha^2 \tag{7}$$

其中：$\Lambda=diag\{\lambda_1,\cdots,\lambda_l\}$；$T_\alpha^2$ 表示置信度为 α 的控制限。设过程正常运行时的样本服从多元正态分布，T_α^2 为

$$T_\alpha^2=\frac{l(n^2-1)}{n(n-l)}F_{l,n-l;\alpha} \tag{8}$$

其中：$F_{l,n-l;\alpha}$ 是带有 l 和 n-l 个自由度、置信度为 α 的 F 分布的临界值。

在计算出每个模态统计量的控制限后，为了使多模态数据进行统一的检测，需要对各个模态的控制限进行匹配。本文以选取公倍数的方法确定每个模态控制限的匹配系数，从而达到统一控制限的目的[18]。多模态 SPE 统计量统一的控制限公式如下：

$$SPE=\prod_{r=1}^{N}SPE_{r,\alpha} \tag{9}$$

其中：N 为模态的总数；模态 r 的 SPE 匹配系数为

$$SPE_r=\prod_{\substack{xr=1\\xr\ne r}}^{N}SPE_{xr,\alpha} \tag{10}$$

多模态 T^2 统计量统一的控制限为

$$T^2=\prod_{r=1}^{N}T_{r,\alpha}^2 \tag{11}$$

模态 r 的 T^2 匹配系数为

$$T_r^2=\prod_{\substack{xr=1\\xr\neq r}}^{N}T_{xr,\alpha}^2 \tag{12}$$

计算新来测试样本与训练样本的概率密度，运用概率密度最大的训练数据标签进行模态划分，计算所属模态标签为 r 下的统计量 $SPE_{r,\alpha}$ 和 $T_{r,\alpha}^2$，再分别乘以匹配系数 SPE_r 和 T_r^2，计算出统一的统计量，从而达到在统一的控制限下进行多模态过程检测的目的。

1.3　建模过程与在线监测

采集正常工况下各个模态的数据，首先对各个模态的训练数据建立 PCA 模型，计算各个模型的控制限和匹配系数。根据匹配系数计算各个模态统一的统计量和控制限。对新来的数据，计算与训练数据的概率密度，运用概率密度最大的训练数据标签确定新数据的模态。将新来数据向对应模态的 PCA 模型上投影，根据匹配系数计算统一的统计量，比较统计量与控制限进行多模态故障检测。

基于 PD-PCA 方法的多模态过程故障检测流程如图 1 所示。其主要分为建模过程与在线监视两部分。

1.3.1 建模过程

1）对正常工况下各个模态的训练数据分别进行标准化处理。

2）建立各个模态的 PCA 模型，根据累积贡献率选择合适的主元个数。

3）计算各个模型的控制限和匹配系数，根据匹配系数计算各个模态统一的统计量和控制限。

1.3.2 在线监视

1）计算新来测试数据与训练数据的概率密度。

2）根据概率密度最大的训练数据标签确定测试数据的模态。

3）将新来数据向对应模态的 PCA 模型上投影，根据匹配系数计算统一的统计量。

4）将统计量与控制限进行比较。若统计量大于控制限，则该反映时刻的数据样本是故障的；否则是正常的。

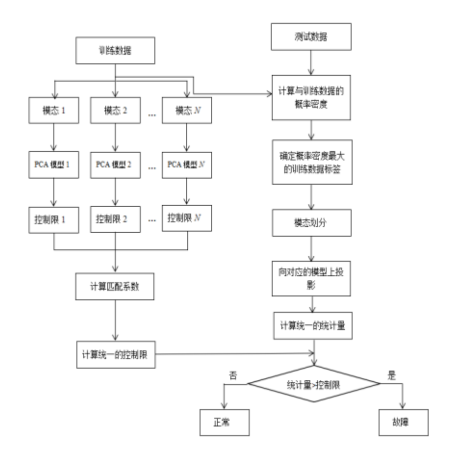


图 1　基于 PD-PCA 方法的多模态过程故障检测流程

2　仿真结果与分析

2.1　工业过程故障诊断

近年来，随着现代工业和科技的进步以及计算机水平的不断提高，化工生产过程设备越来越趋近于复杂化，自动控制系统的规模也越来越大，这不但大大提高了生产效率，同时也降低了生产成本。但同时大规模的化工设备使得影响系统正常运行的因素大大增加，由于系统内部的关联越来越密切，一个部件的异常就有可能导致整个自动控制系统故障。另外，化工行业生产环境条件非常恶劣，一旦操作不当，疏于检查或者出现自然灾害，就有可能出现爆炸、毒气泄漏的危险。目前，在化工生产过程中，爆炸或泄露现象时有发生，这不仅使生产遭受巨大损失，有时甚至会造成人员伤亡，甚至对生态环境造成恶劣的影响。因此，保证化工生产过程的安全稳定运行尤为重要。故障诊断技术作为一种监测和诊断生产过程故障的有效方法，对于提高现代自动控制系统的可靠安全运行、保证化工过程的有效操作具有重要作用。本文通过基于概率密度的 PCA 算法及时发现故障，对保证化工生产过程的正常运行具有重要意义。

2.2　多模态数值例子

本节通过一个人工合成的多模态数值例子[19]，验证基于概率密度的 PCA 算法对多模态过程故障检测的有效性。在此数据中每个样本有两个变量，并且线性相关。在模态 1 中，变量 y_1 服从[0，1]的均匀分布；在模态 2 中，变量 y_1 服从[1，2]的均匀分布。在 2 个模态中，*noise* 服从均值为 0、标准差为 0.01 的正态分布的噪声。具体数据来源于式(13)：

$$\begin{aligned}&\text{Mode 1:}\quad y_2=2y_1+noise\\&\text{Mode2:}\ y_2=1.5y_1+noise+6\end{aligned} \tag{13}$$

按上述模型产生两个模态的 500 个正常数据作为训练样本，100 个正常数据作为校验样本，5 个异常数据作为故障样本。图 2 是训练样本、校验样本、故障样本的数据分布散点图。从图 2 可以看出该数值例子是多模态的。

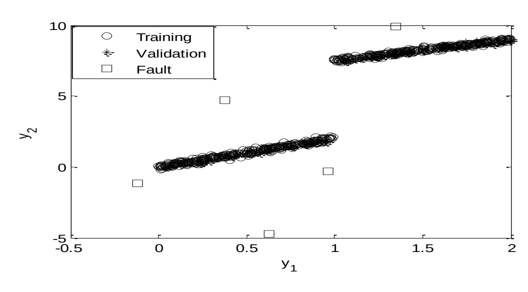


图 2　数据分布散点图

通常所用的故障检测性能指标，有故障检测率、误报率和漏报率等。故障检测率是指系统诊断报告出实际发生故障的部件或设备的数量占所有发生故障部件的百分比。即检测到超出控制限的故障样本占所有故障样本的百分比。误报是指实际没有故障情况而系统诊断报告出某一部件或设备故障。误报率定义为超出控制限的正常样本占所有正常样本的百分比。漏报是

指故障已经发生，而系统未能诊断到作出报警响应或指示等处理。漏报率定义为未超出控制限的故障样本占所有故障样本的百分比。故障检测率和漏报率之和为 1。文中选择误报率和漏报率作为指标，评价算法的故障检测效果。

本文对多模态数值例子运用 PCA、KNN、LOF 和 PD-PCA 方法进行故障检测，检测结果如图 3 所示。KNN 中，近邻数 k=3。LOF 中，近邻数为 23。PCA 和 PD-PCA 的主元个数通过累计贡献率确定。从图 3 可以看出，PCA 算法的 SPE 统计量检测出 1 个故障数据，2 个样本出现误报；T^2 统计量检测出 1 个故障样本，3 个样本出现误报。KNN 算法检测出个 4 故障数据，4 个样本出现误报。LOF 算法检测出 1 个故障数据，2 个样本出现误报。PD-PCA 算法的 SPE 统计量检测出全部的故障数据，无样本出现误报；T^2 统计量检测出 4 个故障样本，无样本出现误报。与 PCA、KNN 和 LOF 方法相比，PD-PCA 故障检测率较高，误报率相对较低，从而验证 PD-PCA 算法的有效性。

表 1 是四种算法对数值例子的检测结果。由表 1 可以看出，PD-PCA 算法的 SPE 检测指标和 T^2 检测指标检测效果均明显优于 PCA 的检测效果。KNN 算法的漏报率较低，但是相对 PD-PCA 算法其误报率相对较高。LOF 算法的误报率较低，但是未完全检测出全部的故障，这是由于 LOF 易受离群样本的影响。综上所述，与其他三种算法对比，PD-PCA 算法有较低的误报率和漏报率，验证了该算法的有效性和优越性。

表 1　多模态数值例子的检测结果

	PCA		KNN	LOF	PD-PCA	
	SPE	T^2	距离	LOF	SPE	T^2
误报率(%)	2	3	4	2	0	0
漏报率(%)	80	40	20	80	0	20

2.3 半导体生产过程数据

半导体生产过程作为一个完善的工业过程仿真平台，在基于数据驱动的故障检测研究领域得到了广泛的应用。本文应用半导体工业实例——A1 堆腐蚀过程[20,22]比较不同的故障检测方法的性能。半导体数据来源于美国德州仪器公司的半导体生产过程实际数据，是个典型的非线性、时变、多阶段和多工况的复杂多模态间歇过程。数据集中的变量 EndPt A 表明了这些特性，如图 4 所示。该数据是由 3 个模态的 108 个正常批次和 21 个故障批次组成, 这些数据分别从 3 次实验（1996 年二月、三月和四月）中产生的。由于两个批次过程（第 56 个正常批次和第 12 个故障批次）丢失大量的数据，所以实际的批次为 107 批正常数据和 20 批故障数据。通过改变变量产生 20 个故障如表 2 所示。在 107 个正常批次中 1~34 批次为第一模态，35~70 批次为第二模态， 71~107 批次为第三模态。每个模态分别选取 32 个批次用于建模，剩下的正常批次作为校验批次用来验证模型的准确性，因此建模批次为 96 个，正常校验批次为 11 个，故障批次为 20 个。从 21 个测量变量中选取 17 个变量作为检测变量，如表 3 所示。

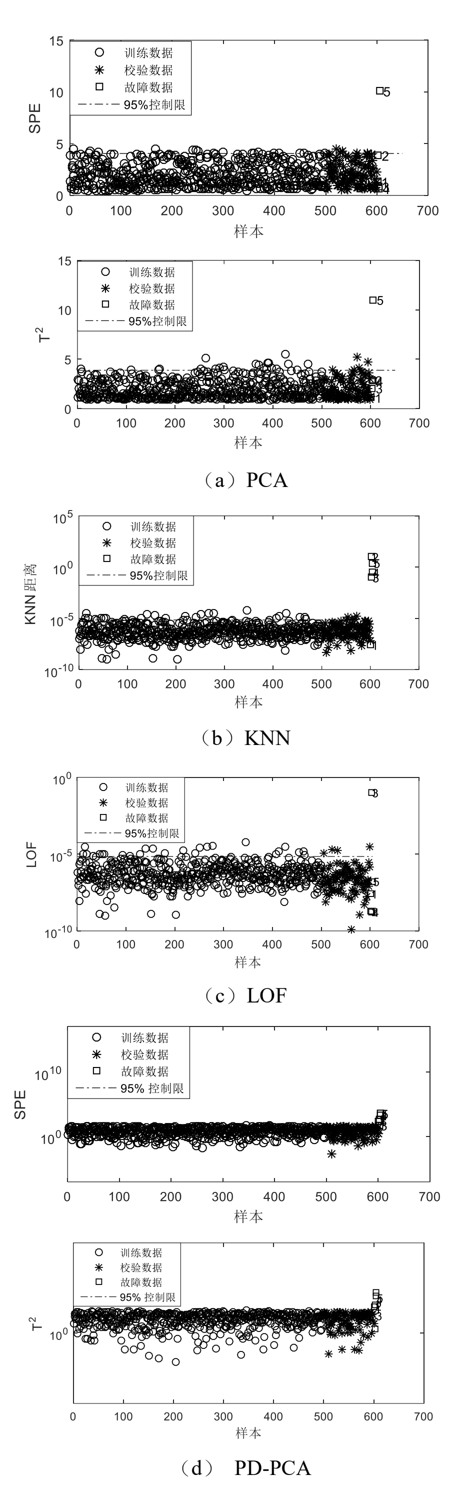


图 3　四种方法对多模态数值例子的故障检测图

每个批次是不等长的，持续时间在 95~112 s 间变化。本文运用最短长度法获得等长批次。为了消除传感器中初始的波动影响，去除开始的 5 个样本，保留 85 个样本以适应最短的批次。将三维建模数据 $\boldsymbol{X}$（96×85×17）沿批次方向展开成二维矩阵 $\boldsymbol{X}$（96×1445）。对校验数据和故障数据也进行同样的处理。

对二维数据矩阵分别运用 PCA、KNN、LOF 和 PD-PCA 方法进行建模，并对 11 个校验批次和 20 个故障批次数据进行故障检测。KNN 中，近邻数 k=3。 LOF 中，近邻数为 23。 PCA 和 PD-PCA 的主元个数由主元贡献率决定。表 4 是四种算法对半导体数据的检测结果。由表 4 可以看出，PCA 和 PD-PCA 算

法的 T^2 检测指标检测出相同数量的故障批次，PD-PCA 算法的误报率相对较高，但是 PD-PCA 算法的 SPE 检测指标检测出全部的故障批次，而且没有误报，其故障检测效果明显优于 PCA 算法。KNN 和 LOF 算法的误报率较低，但是未完全检测出全部的故障，这是由于 KNN 和 LOF 易受离群样本的影响。综上所述，与其他三种算法对比，PD-PCA 算法有较低的误报率和漏报率，说明该方法对于多模态数据的故障检测非常有效，验证了该算法的有效性和优越性。

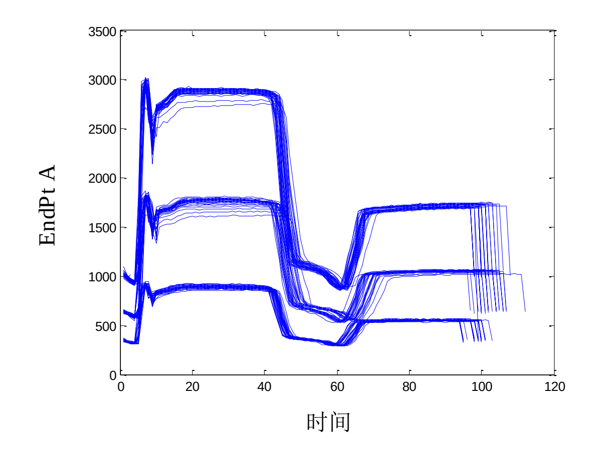


图 4 数据集中的变量 EndPt A

表 2 半导体生产过程所用的故障

序号 序号	故障	序号	故障
1	TCP + 50	11	C12 + 5
2	RF - 12	12	BC13 - 5
3	RF + 10	13	Pr + 2
4	Pr + 3	14	TCP - 20
5	TCP + 10	15	TCP - 15
6	BC13 + 5	16	C12 - 10
7	Pr - 2	17	RF - 12
8	C12 - 5	18	BC13 + 10
9	He Chuck	19	Pr + 1
10	TCP + 30	20	TCP + 20

表 3 半导体生产过程所用检测变量

序号	过程变量	序号	过程变量
1	BC13 流量	10	RF 功率
2	C12 流量	11	RF 阻抗
3	RF 底部功率	12	TCP 调谐
4	A 检测端点	13	TCP 相位误差
5	氦压力	14	TCP 阻抗
6	Helium 压强	15	TCP 顶部功率
7	RF 调谐	16	TCP 负荷
8	RF 负荷	17	Vat 阀门
9	相对误差		

表 4 四种算法对半导体数据的检测结果

故障	PCA		KNN	LOF	PD-PCA	
批次	SPE	T^2	距离	LOF	SPE	T^2
1	√		√	√	√	√
2	√		√		√	
3					√	
4	√	√	√	√	√	
5					√	
6					√	
7	√	√	√	√	√	
8	√		√	√	√	
9	√				√	
10	√		√	√	√	√
11	√		√	√	√	
12	√	√	√	√	√	√
13	√	√	√	√	√	√
14	√		√	√	√	√
15	√		√	√	√	
16	√		√	√	√	
17	√		√	√	√	
18	√	√	√	√	√	
19	√		√	√	√	
20	√		√	√	√	
误报率(%)	0	0	0	0	0	18
漏报率(%)	15	75	20	35	0	75

3 结束语

本文提出一种基于 PD-PCA 的多模态过程故障检测方法。该算法用于处理多模态故障检测过程，能够最大化地分离多模态的正常和故障数据。运用概率密度能够准确的进行模态识别，从而更加准确地检测出校验数据和故障数据来自哪个模态。将本文方法应用到数值例子和实际的半导体工业数据中，仿真结果表明，与传统的故障检测算法相比，本文算法降低了误报率和漏报率，验证了该方法的有效性。

参考文献：

[1] Venkatasubramanian V, Rengaswamy R, Yin K, *et al.* A review of process fault detection and diagnosis: part Ⅰ: Quantitative model-based methods [J]. Computers & Chemical Engineering, 2003, 27 (3): 293-311.

[2] Venkatasubramanian V, Rengaswamy R, Yin K, *et al.* A review of process fault detection and diagnosis: part Ⅱ: Quantitative models and search strategies [J]. Computers & Chemical Engineering, 2003, 27 (3): 313-326.

[3] Venkatasubramanian V, Rengaswamy R, Yin K, *et al.* A review of process fault detection and diagnosis: Part Ⅲ: Process history based methods [J]. Computers & Chemical Engineering, 2003, 27 (3): 327-346.

[4] 周东华，李钢，李元. 数据驱动的工业过程故障检测与诊断技术 [M]. 北京：科学出版社, 2011: 1-76.

[5] 郭金玉，齐蕾蕾，李元. 基于 DMOLPP 的间歇过程在线故障检测 [J]. 仪器仪表学报, 2015, 36 (1): 28-36.

[6] Hu K, Yuan J. Statistical monitoring of fed-batch process using dynamic multiway neighborhood preserving embedding [J]. Chemometrics and intelligent laboratory systems, 2008, 90 (2): 195-203.

[7] Zhao Shijian, Zhang Jie, Xu Yongmao. Performance monitoring of process with multiple operation modes through multiple PLS models [J]. Journal of Process Control, 2006, 16 (7): 763-772.

[8] Yu Jie. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes [J]. Chemical Engineering Science, 2012, 68 (1): 506-519.

[9] He Q P, Wang J. Fault detection using the K-nearest neighbor rule for semiconductor manufacturing processes [J]. IEEE Trans on Semiconductor Manufacturing, 2007, 20 (4): 345-354.

[10] 马贺贺，胡益，侍洪波. 基于马氏距离局部离群因子方法的复杂化工过程故障检测 [J]. 化工学报, 2013, 64 (5): 1674-1682.

[11] 刘帮莉，马玉鑫，侍洪波. 基于局部密度估计的多模态过程故障检测 [J]. 化工学报, 2014, 65 (8): 3071-3081.

[12] 李元，马雨含，郭金玉. 基于动态多向局部离群因子的在线故障检测 [J]. 计算机应用研究, 2017, 34 (11): 3259-3261, 3266.

[13] Deng Xiaogang, Tian Xuemin. Multimode process fault detection using local neighborhood similarity analysis [J]. 中国化学工程学报：英文版, 2014, 22 (11-12): 1260-1267.

[14] 郭金玉，陈海彬，李元. 基于主样本模型的间歇过程 KNN 故障检测方法 [J]，信息与控制, 2014, 43 (4): 495-500.

[15] Jiang Q, Yan X, Tong C. Double-weighted independent component analysis for non-Gaussian chemical process monitoring [J]. Industrial & Engineering Chemistry Research, 2013, 52 (40): 14396-14405.

[16] Rosenblatt M. Curve estimates [J]. Annals of Mathematical Statistics, 1971, 42 (6): 1815-1842.

[17] Parzen E. On estimation of a probability density function and mode [J]. Annals of Mathematical Statistics, 1962, 33 (3): 1065-1076.

[18] 郭金玉，袁堂明，李元. 一种不等长的多模态间歇过程故障检测方法 [J]. 化工学报, 2016, 67 (7): 2916-2924.

[19] 谭帅，王福利，常玉清，等. 基于差分分段 PCA 的多模态过程故障检测 [J]. 自动化学报, 2010, 11 (36): 1626-1636.

[20] Wise B M, Gallagher N B, Butler S W, *et al.* A comparison of principal component analysis, multiway principal component analysis, trilinear decomposition and parallel factor analysis for fault detection in a semiconductor etch process [J]. Journal of Chemometrics, 2015, 13 (3-4): 379-396.

[21] Lee S P, Chao A K, Tsung F, *et al.* Monitoring batch processes with multiple on-off steps in semiconductor manufacturing [J]. Journal of Quality Technology, 2011, 43 (2): 142-157.

[22] Yu J. Fault detection using principal components based Gaussian mixture model for semiconductor manufacturing processes [J]. Semiconductor Manufacturing, 2011, 24 (3): 432-444.